

University of Groningen

Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours

Ribeiro, Cássia O; Knopf, Antje; Langendijk, Johannes A; Weber, Damien C; Lomax, Antony J; Zhang, Ye

Published in:
Radiotherapy and Oncology

DOI:
[10.1016/j.radonc.2018.03.001](https://doi.org/10.1016/j.radonc.2018.03.001)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2018

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Ribeiro, C. O., Knopf, A., Langendijk, J. A., Weber, D. C., Lomax, A. J., & Zhang, Y. (2018). Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours. *Radiotherapy and Oncology*, 128(1), 174-181.
<https://doi.org/10.1016/j.radonc.2018.03.001>

Copyright

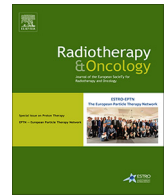
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.



Proton therapy

Assessment of dosimetric errors induced by deformable image registration methods in 4D pencil beam scanned proton treatment planning for liver tumours

Cássia O. Ribeiro^{a,*}, Antje Knopf^a, Johannes A. Langendijk^a, Damien C. Weber^{b,c}, Antony J. Lomax^{b,d}, Ye Zhang^b^a Department of Radiation Oncology, University Medical Center Groningen, University of Groningen, The Netherlands; ^b Center for Proton Therapy, Paul Scherrer Institut, Villigen-PSI; ^c Radiation Oncology Department, University Hospital Zurich; and ^d Department of Physics, ETH Zurich, Switzerland

ARTICLE INFO

Article history:

Received 5 September 2017
Received in revised form 16 February 2018
Accepted 6 March 2018
Available online 20 March 2018

Keywords:

Liver cancer
Pencil beam scanned proton therapy
Deformable image registration
Deformation vector field
4DCT-MRI data sets
4D dose calculation

ABSTRACT

Purpose: Respiratory impacts in pencil beam scanned proton therapy (PBS-PT) are accounted by extensive 4D dose calculations, where deformable image registration (DIR) is necessary for estimating deformation vector fields (DVF). We aim here to evaluate the dosimetric errors induced by different DIR algorithms in their resulting 4D dose calculations by using ground truth (GT)-DVF from 4DMRI.**Materials and methods:** Six DIR methods: ANACONDA, Morfeus, B-splines, Demons, CT Deformable, and Total Variation, were respectively applied to nine 4DCT-MRI liver data sets. The derived DVFs were then used as input for 4D dose calculation. The DIR induced dosimetric error was assessed by individually comparing the resultant 4D dose distributions to those obtained with GT-DVFs. Both single-/three-field plans and single/rescanned strategies were investigated.**Results:** Differences in 4D dose distributions among different DIR algorithms, and compared to the results using GT-DVFs, were pronounced. Up to 40 % of clinically relevant dose calculation points showed dose differences of 10 % or more between the GT. Differences in $V_{95}(\text{CTV})$ reached up to 11.34 ± 12.57 %. The dosimetric errors became in general less substantial when applying multiple-field plans or using rescanning.**Conclusion:** Intrinsic geometric errors by DIR can influence the clinical evaluation of liver 4D PBS-PT plans. We recommend the use of an error bar for correctly interpreting individual 4D dose distributions.© 2018 The Author(s). Published by Elsevier B.V. Radiotherapy and Oncology 128 (2018) 174–181 This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

When treating moving targets in the thorax or abdomen with pencil beam scanned proton therapy (PBS-PT), due to the presence of breathing motion, a time-resolved 3D (4D) image is necessary for quantifying the motion characteristics and performing a 4D dose calculation. For treating this type of tumour with a highly precise technique such as PBS-PT, a 4D dose calculation is crucial in order to take into account the deterioration of the dose distribution due to the relative motion between the target and the delivered pencil beams (interplay effects) [1–3].

To calculate motion induced geometric differences between two image phases, deformable image registration (DIR) is the standard approach for building up a point-to-point correlation between corresponding features. To perform DIR, a fixed and a moving image are pre-defined to estimate the patient's deformable motion between these two images [4]. The result of DIR is a deformation

vector field (DVF), which contains vectors for each voxel pointing from the fixed image towards the moving image.

For any form of radiotherapy, DIR is one of the irreplaceable components for both inter- and intra- fractional dosimetric evaluation. It is especially important for PBS-PT, due to its high sensitivity to geometric accuracy. However, it is well known that DIR is an ill-posed problem intrinsically [5]. When applying different DIR methods to the same image pair, the resulting motion estimations can be inaccurate and differ significantly from each other [6]. Some of these errors are quantifiable, and can be calculated by comparing the DIR estimated motion of well-defined landmarks to their actual positions in both images (the so-called ground truth (GT) data). This is the classic approach of evaluating any DIR algorithm performance, as used by many previous publications [7,8]. Despite compromising the efficiency for the error quantification, the more landmarks that are defined, the more reliable the results will be. In contrast, there are also unquantifiable errors in featureless regions of the images, where the deformable problem is intrinsically ill-defined. Motion vectors in these regions will directly depend on

* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands.
E-mail address: c.oraboni.ribeiro@umcg.nl (C.O. Ribeiro).

the transformation model and regularization of the selected algorithm, and it is therefore unavoidable that ambiguity between different algorithms will exist.

Any form of registration uncertainty can directly lead to distinguishable differences in dose distributions, which consequently influence any further dosimetric analysis and clinical decision-making [9]. In the literature, a number of studies have investigated the dosimetric uncertainties induced by a particular DIR method [10,11]. However, their conclusions were restricted to their selected DIR method, and a consensus on the clinical impact of DIR uncertainty is still difficult to achieve. Yeo et al. [12] compared calculated doses based on results from several available DIR algorithms with a measured dose using a deformable 3D dosimeter. However, DIR errors for real patient geometries may perform differently in contrast to the rather simple experimental setup used in that work.

Zhang et al. [6] showed that the ambiguity of two DIR approaches can lead to significant differences in the estimated motion maps, and subsequent 4D dose distributions, among liver cancer patients for PBS-PT, even if landmark registration errors were similar. Due to the lack of a comprehensive GT-DVF however, it is often impossible to validate the accuracy of DIR in the whole region of interest.

In this work, we would like to improve the above studies in two aspects. First, to investigate the systematic errors induced by DIRs in 4D dose calculations, using the unique advantage of comprehensive GT-DVFs extracted from synthetic 4DCT-MRI [13]. Second, we include multiple DIR methods to reveal the extent of potential variation induced by different algorithms. As such, six DIR methods (five commercially available and one research version) have been applied to nine 4DCT-MRI data sets to estimate deformable motion within the abdomen region. Compared to previous works, we also consider comprehensive GT-DVFs as reference to quantify the absolute accuracy for deformable motion estimation. Consequently, the resulting 4D dose distributions generated using different DIR algorithms can be directly compared under conditions of varying plan configurations, rescanning scenarios, patient geometries, and motion scenarios.

Materials and methods

Synthetic 4DCT-MRI and ground truth deformation vector fields (GT-DVFs)

4DCT-MRI data sets consist of end-of-exhalation 3DCTs (reference phases) modulated by consecutive and extended breathing motion extracted from 4DMRI data through a validated image processing method [13–15] (Fig. 1, upper left). Through this process, synthetic 4DCT-MRI data sets within the liver are obtained by warping the reference phase with DVFs extracted from 4DMRI using a combination of multiresolution affine registration and B-spline non-rigid registration [15].

Nine such 4DCT-MRI data sets, generated from motion artefact-free 3DCTs of three liver cancer patients (denoted as PI, PII, and PIII respectively), were included in this study. The reference phases of the three patients were modulated by three different 4DMRI motion scenarios indicated as motions A, B, and C [16]. Clinical target volumes (CTVs) at the reference phase were 122, 264, and 403 cm³ for patients I, II, and III respectively. Only 4DCT-MRI data sets corresponding to the first breathing cycle were analysed, and no consideration of motion irregularity has been included in the study. For the nine data sets, the amplitude for the first breathing cycle (given by the mean of the amplitude of all different points within the whole liver region) of motion scenarios A, B, and C were 7.82 (SD = 2.01), 20.61 (SD = 3.39), and 16.88 (SD = 2.78) mm respectively. Additionally, motion periods (extracted using Fourier

analysis) for this first cycle equalled 3.66, 4.62, and 7.22 s for A, B, and C respectively. The corresponding DVFs extracted from 4DMRI to generate these nine 4DCT-MRI data sets were then defined as the GT-DVFs. Subsequently, new DVFs were extracted from these nine 4DCT-MRI data sets using the different DIR methods being investigated (see Fig. 1). These GT-DVFs and DIR estimated DVFs were used for the 4D dose calculation analysis.

Deformable image registration (DIR) methods and derived deformation vector fields (DVFs)

Six DIR methods have been included in this study. DIR1 and DIR2 are available in the RayStation (RaySearch Laboratories, Stockholm, Sweden) treatment planning system used in the UMCG, whereas DIR3 and DIR4 [6] are algorithms provided in open source software (Plastimatch; www.plastimatch.com) and used at PSI. DIR5 and DIR6 were developed in turn by the commercial medical imaging software company Mirada Medical (Oxford, UK) and by the Computer Vision Laboratory in ETH Zurich (Zurich, Switzerland) respectively. The different DIR methods are based on the ANACONDA [17], Morfeus [18,19], B-splines, Demons, CT Deformable [20,21], and Total Variation [22] algorithm respectively (Suppl. 1). For each data set, all six approaches were applied to the reference phase as the fixed image. The remaining phases were defined as successive moving images (see Table S.1).

4D dose calculation

The DVFs resulting from the application of the six DIR methods were used as input to the in-house 4D dose calculation engine at PSI, which is an extension of the 3D dose calculation algorithm. The gantry (beam) coordinate system is defined as (s, t, u) , in which s is the pencil beam central axis direction and (t_0, u_0) its position orthogonal to the field direction (Fig. 2(a)). The clinically used dose grid size in this coordinate system is $4 \times 4 \times 2.5$ mm³.

To extend the 3D dose calculation to a 4D dose calculation, time-dependent displacements of dose grid points for motions in the t and u directions are taken into account using displacement and density-variation maps derived from each phase of the relevant 4DCT-MRI data. The 4D dose calculation algorithm first estimates the time stamp of each delivered pencil beam [13]. The DIR extracted DVFs are then geometrically translated and rotated into the gantry (s, t, u) coordinate system, and sampled by the dose grid size to provide displacement maps for each dose grid point [6]. Density-variation maps are derived from the different 4DCT-MRI phases using Siddon's algorithm [23]. With these displacement and density-variation maps, the offsets of the dose grid points from their nominal positions are calculated and a 4D dose distribution obtained.

4D planning configurations

Static, single-field uniform dose (SFUD) plans [24] were calculated on each of the reference phases of the three patients. Both single- and three-field plans were investigated. Field arrangements were anterior-posterior (F1), right lateral (F2), and anterior-inferior oblique (F3), with the three-field plan being a combination of all fields. 4D dose distributions for these were then subsequently obtained by using either the GT-DVFs or the DVFs resulting from the six DIR methods in the 4D dose calculation algorithm (Fig. 1). Single scan or five times layered rescanning [16] were simulated with the scanning parameters of Gantry 2 at PSI [25–27]. Plan delivery started at the reference phase of the corresponding 4DCT-MRI. All analysed plan configurations and respective notations are given in Suppl. 2.

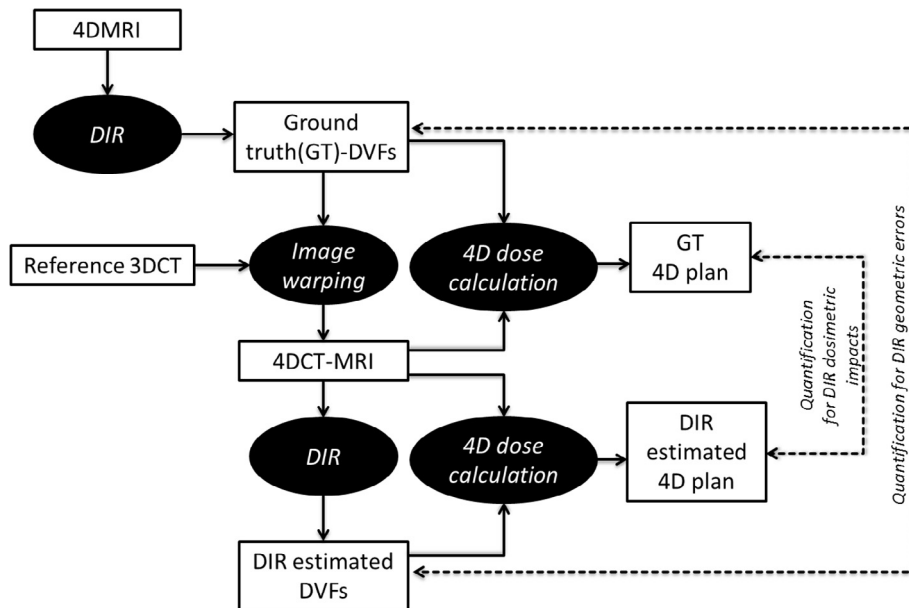


Fig. 1. Schematic representation of the workflow adopted in this paper.

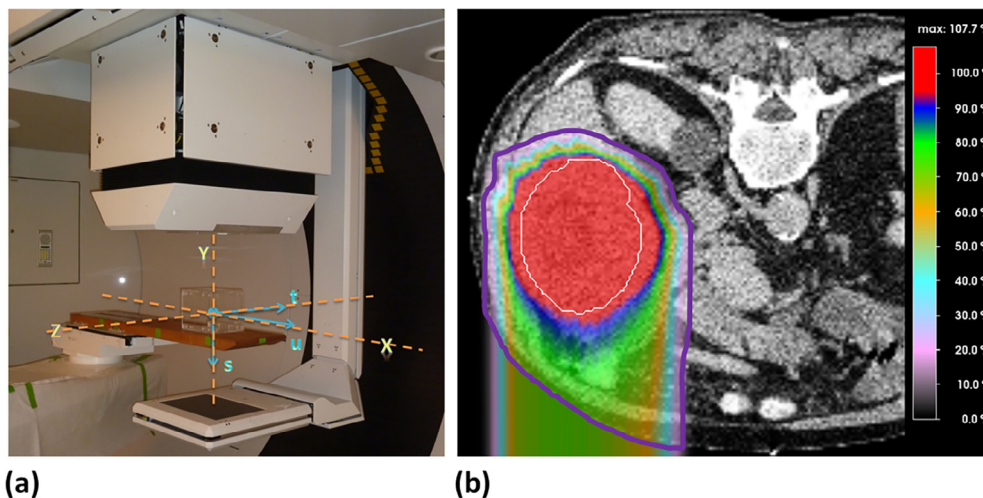


Fig. 2. (a) Gantry (or beam) coordinate system defined by (s, t, u) represented in blue. (b) Anterior-posterior (F1) fVOI of patient PIII (in purple) for the corresponding geometric field-specific registration error quantification.

Assessing DIR performance

The impact of using the different DVFs derived from the applied DIR methods in the 4D dose calculation has been assessed by (1) quantifying the geometric error in the beam coordinate system and (2) analysing dosimetric errors in the 4D dose distribution.

The DIR induced geometric error was quantified by the absolute differences of the derived DVFs from each DIR method with respect to the GT-DVFs (Suppl. 3). Since this study is focused on the impact of DIR on 4D dose distributions, the assessment of the geometric error was performed in the field direction, by analysing the t and u coordinates within a defined field-specific volume of interest (fVOI) (Fig. 2(b)). To more precisely correlate the geometric error to the dosimetric error, we defined the fVOI as the dose region where the dose calculation grid covers patient geometry (to exclude dose regions outside the patient). As such, the geometric evaluation was performed only in regions that directly contribute to the 4D dose calculation.

The 4D dose distributions obtained from the DVFs derived from the six DIRs were individually calculated and compared to those resulting when using the GT-DVFs. Both dose-volume histograms (DVHs) with error bands, as well as $V_{95}(\text{CTV})$ values were analysed to quantify the impact of DIR in clinical practice. Difference DVHs (DDVHs), which correspond to the histograms of the absolute dose difference between the individual DIR estimated and GT 4D plans in an extension of the CTV volume ($\text{CTV} + 1 \text{ cm}$), were computed. Finally, for each investigated DIR method, percentages of the extended CTV volume with absolute dose differences higher than 10 % were extracted from the DDVH for analysis.

Results

DIR induced geometric error

Geometric registration errors with respect to motion scenarios and DIR methods for all single-field arrangements are given in

Table 1

Field-specific geometric errors before and after DIR for the analysed configurations: three different single-field directions and three liver cancer patients, each modulated by motion scenarios A, B, or C.

Motion	Before DIR (mean \pm SD) [mm]	After DIR (mean \pm SD) [mm]						
		GT	DIR1	DIR2	DIR3	DIR4	DIR5	DIR6
A	1.47 \pm 0.26		1.05 \pm 0.29	1.01 \pm 0.15	1.04 \pm 0.23	0.76 \pm 0.15	0.74 \pm 0.17	0.72 \pm 0.16
B	3.91 \pm 0.57		2.44 \pm 0.88	2.34 \pm 0.33	2.76 \pm 0.56	1.97 \pm 0.46	1.93 \pm 0.52	1.85 \pm 0.55
C	4.47 \pm 0.57		2.84 \pm 0.52	2.66 \pm 0.35	3.21 \pm 0.54	2.38 \pm 0.35	2.34 \pm 0.38	2.22 \pm 0.42
								Averaged
								0.89 \pm 0.06
								2.22 \pm 0.18
								2.61 \pm 0.08

Table 1. The lowest and highest mean errors in the field direction for the smallest motion cases (motion scenario A) were achieved by DIR6 and DIR1 (0.72 ± 0.16 mm and 1.05 ± 0.29 mm respectively). Despite motion B having the highest amplitude within the liver (Suppl. 4.1), motion C actually showed larger movement in the different single-field directions within the selected fVOI. For motion C, all the errors increased up to 3.21 ± 0.54 mm for DIR3, 2.84 ± 0.52 mm for DIR1, and 2.22 ± 0.42 mm for DIR6. Overall, the best performance was achieved by DIR6 for all three motion scenarios. Most importantly, it was also observed that all DIR methods underestimated the GT motion amplitudes (Table S.4.1).

DIR induced dosimetric error

Without rescanning, pronounced differences in 4D dose distributions were observed among the different DIR scenarios, and in comparison to the GT generated 4D dose distributions (Fig. 3(a)). Indeed, the impact of interplay was clearly different when using different DVFs as input to the 4D dose calculation. For the GT, DIR2, DIR4, DIR5, and DIR6 the effects of interplay were more pronounced than for the other two methods (DIR1 and DIR3), in which the dose distributions look more homogeneous. In fact, for this particular 4D plan configuration, DVHs of the CTV and DDVHs of the extended CTV volume, obtained from the six DIR 4D plans, clearly differed from the GT DVH (Fig. 3(b.i) and (b.ii)), with all DIR plans underestimating the dose in-homogeneity due to the interplay effects in comparison to the GT plan. Additionally, substantial absolute $V_{95}(\text{CTV})$ differences (between the GT and all investigated DIR methods) were observed (Fig. 4). For motion A, and single fields with single scans, $V_{95}(\text{CTV})$ differences of $7.91 \pm 3.46\%$ were observed for DIR3, and $2.02 \pm 1.28\%$ for DIR2. Moderate motions in the fVOI (motion scenario B) and single fields without rescanning showed the greatest differences, with DIR3 and DIR2 having differences of $10.58 \pm 14.08\%$ and $1.43 \pm 1.37\%$ respectively. Of the six tested DIR methods, DIR2 achieved the lowest errors in $V_{95}(\text{CTV})$ for most situations with the best prediction of interplay effects in comparison to the GT.

Rescanning and/or multiple-field plans smoothed out the $V_{95}(\text{CTV})$ differences for all three motion scenarios. For instance, for motion scenario B, three-field plans with rescanning resulted in absolute $V_{95}(\text{CTV})$ differences to the GT of $3.46 \pm 1.40\%$ and $0.23 \pm 0.19\%$ for DIR3 and DIR2 respectively. For the smallest motion, negligible dosimetric errors were obtained for all DIR methods ($0.37 \pm 0.38\%$ for DIR3, $0.30 \pm 0.17\%$ for DIR2, and $0.24 \pm 0.18\%$ for DIR4). $V_{95}(\text{CTV})$ values for all 4D plan configurations can be found in Table S.4.2, showing that overestimation of target coverage is consistent for the single-field single/rescanned, or multiple-field single scan DIR generated plans.

DDVH bands in the extended CTV for all 4D plan configurations (as a function of DIR for each motion scenario (Figs. S.4.2.1 and S.4.2.2)) show that the accuracy of a particular DIR method does not depend just on the motion characteristics, or that any single DIR algorithm performed better for one particular motion scenario. Generally however, dose differences between DIR2 and the GT were the smallest for most of the 4D plan configurations and for

all analysed motion scenarios. For the others, up to 40 % of the volume of extended CTV had absolute dose differences in comparison to the GT of more than 10 % for single-field plans delivered with a single scan (Fig. 5(a)). The largest dosimetric differences occurred for single-field plans delivered with a single scan for the largest motions (B and C). Dose differences using all DIRs could however be smoothed out when applying rescanning or adding fields to the treatment plan (Fig. 5(b)).

Discussion

We have shown in this study that the application of different DIR methods to extract motion from 4D data sets can result in pronounced differences. Geometric differences of up to 1.05 ± 0.29 mm for the smallest motion amplitude (motion A) and 3.21 ± 0.54 mm for larger motions (motion C) have been observed. Without motion mitigation, associated dosimetric differences in target coverage (assessed as differences in $V_{95}(\text{CTV})$) were as high as $7.91 \pm 3.46\%$ for motion A and $11.34 \pm 12.57\%$ for motion B. Moreover, up to 40 % of the extended CTV volumes have been predicted to have absolute dose differences in comparison to the GT dose of more than 10 %. However, differences in 4D dose distributions among different DIR scenarios, and compared to the GT, were significantly smoothed out in most cases when using multiple-field treatment plans and/or rescanning [26]. In addition, it was confirmed that larger motion amplitudes and deformations contribute to larger geometric registration errors and consequently larger dosimetric errors [6]. However, target coverage (quantified by the $V_{95}(\text{CTV})$) was highly dependent on other factors, such as magnitude of the interplay effects, and no linear relationship between target coverage and motion amplitude could be found.

The 4D dose calculation algorithm selected for this project is the one developed at PSI. It obtains a 4D dose distribution by deforming the dose grid as a function of time, instead of performing multiple calculations of the dose on different 3DCT phases, which need to be subsequently warped back to a reference phase for dose accumulation [28–30]. Thus, it has the potential advantage of allowing for a high number of recalculations in an acceptable time-frame. Additionally, unlike the more standard 4D algorithm, the 4D dose calculation of PSI employs linear interpolation for the present motion between the phases of the 4D imaging and this approach has been shown to provide more accurate results [14,31,32].

Instead of focusing solely on the assessment of registration ambiguity as in the work performed by Zhang et al. [6], this study systematically investigates the performance of a variety of DIR algorithms, with respect to GT-DVFs, as part of a 4D dose calculation. Naturally, such GT-DVFs have also been extracted from 4DMRI using DIR and one could argue therefore, that DIR induced error is already present in these GT-DVFs. As such, these do not represent ‘real’ GT anatomical motion, but rather an estimation that will depend on the characteristics of the extraction method used. Indeed, we cannot deny that DIR induced error is present in the GTs. However, these are treated as GTs for the 4DCT-MRI data set itself (where the different DIRs are applied), and not for

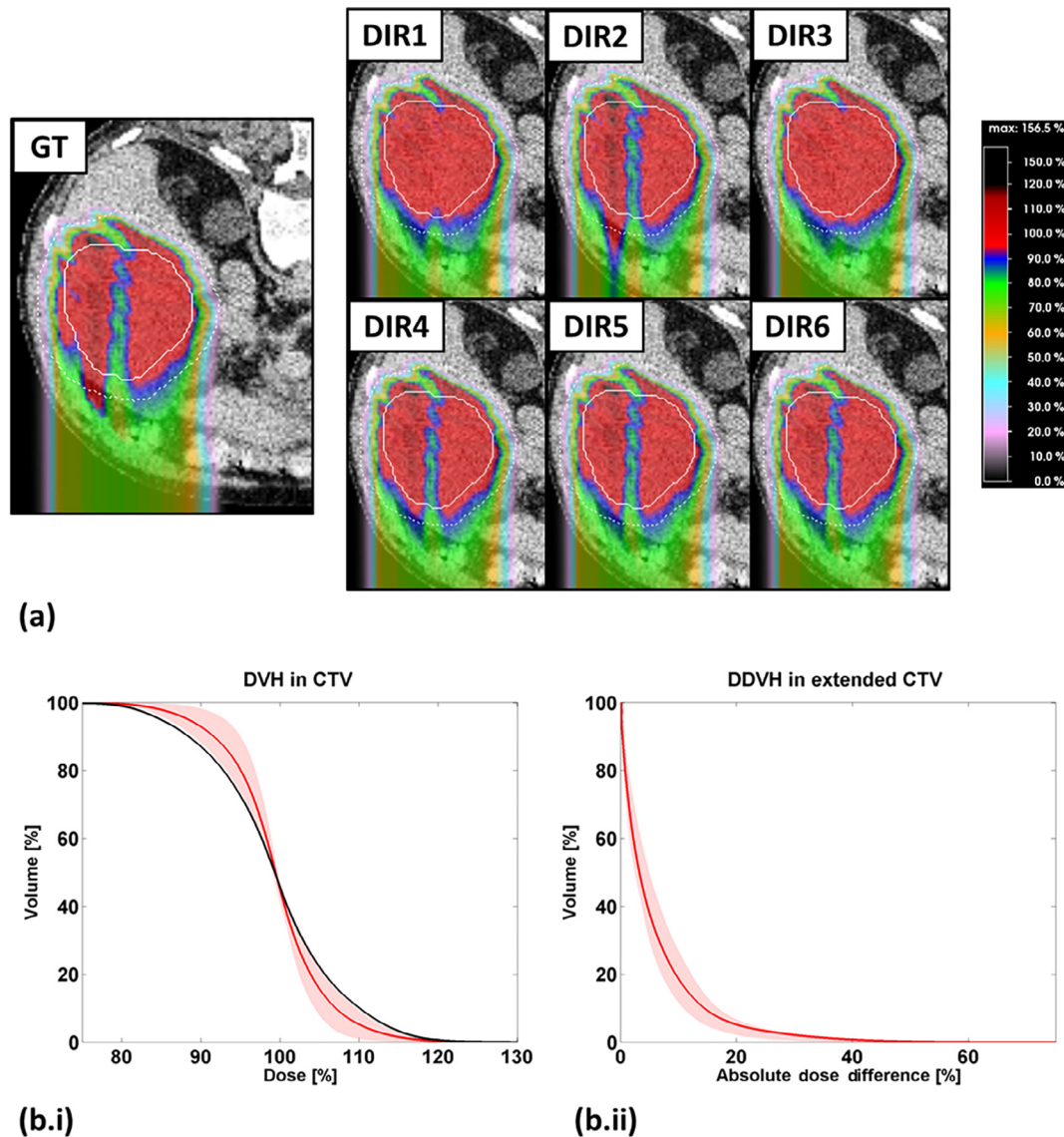


Fig. 3. 4D dose calculation results for the example 4D plan configuration of patient geometry PIII modulated by motion scenario C, and treated with the single anterior-posterior field F1 applied without any rescanning. This corresponds to the patient case with the largest tumour volume and moderate motion amplitude within the liver, but largest movement in the field direction within the selected fVOI. **(a)** 4D dose distributions using GT, DIR1, DIR2, DIR3, DIR4, DIR5, and DIR6 DVFs. The white normal and white dashed lines represent the CTV and CTV + 1 cm (extended CTV) delineations respectively. **(b.i)** CTV DVH and **(b.ii)** extended CTV DDVH curves obtained with the six DIR methods. The black solid line gives the DVH curve calculated with the GT. The red shadow represents the band obtained by using different motion estimation methods, and the solid red line is showing the mean value from the six DIRs.

the 4DMRI, and so these errors turn out to have minimal impact on our conclusions. Additionally, due to the much higher contrast in abdomen MRI images, we believe the DIR errors for the GTs are rather limited. Therefore, for DIR error assessment applications, the GT-DVFs used here as reference provide dense image features for comparison, being a great advantage over the common approach that relies on the sparse distribution of identifiable landmarks [7,8]. Moreover, the recently published American Association of Physicists in Medicine task group 132 report provided recommendations for clinical DIR quality assurance, such as a digital phantom [32].

Although several 4D treatment plan configurations using SFUD have been analysed, further research still needs to be performed analysing 4D intensity-modulated proton therapy (IMPT) plan configurations. In addition, statistics could be improved by setting different starting phases for the treatment delivery, and only one

breathing cycle has been considered, which in the 4D dose calculation was repeated, cycle-to-cycle, over the duration of the full delivery time per field. Results may vary even more if true, variable breathing patterns were taken into account. As such, the influence of irregular breathing patterns, as provided by the 4DCT-MRI approach, will be exploited in future work.

Interestingly, the geometric and dosimetric accuracies provided by all tested DIR methods are not directly correlated. For most situations, Total Variation (DIR6) provided the lowest field-specific geometric errors and Morfeus (DIR2) the lowest dosimetric errors. However, it is important to remember the complex nature of 4D dose calculations, and so multiple other variables could have influenced this lack of correlation. Furthermore, DIR2 is the only algorithm that requires a contour of the liver to be delineated on all the registered image pairs, indicating that the addition of such anatomical information to the algorithm is a safe approach to

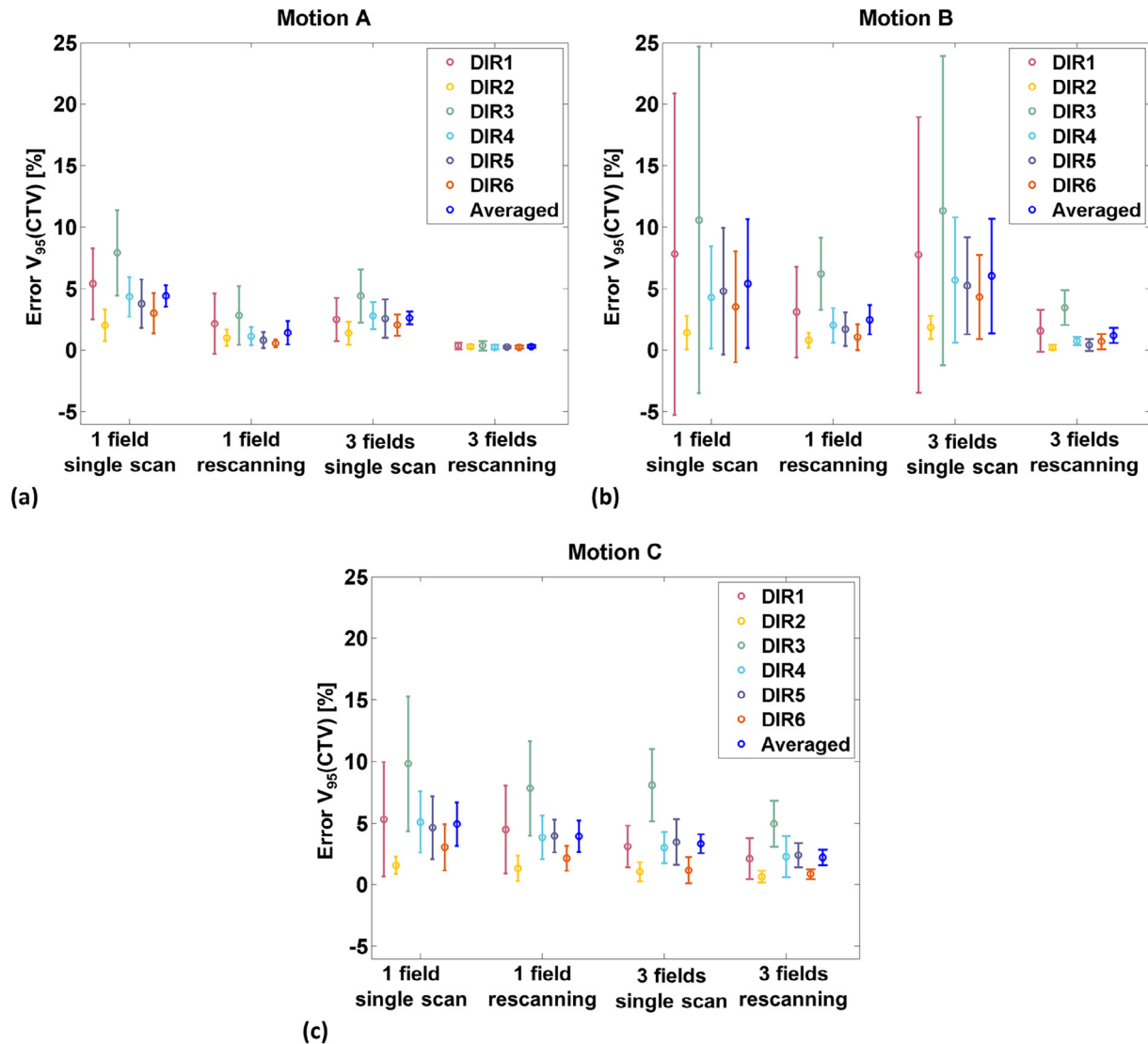


Fig. 4. Absolute differences between the $V_{95}(\text{CTV})$ of a particular DIR method and the GT $V_{95}(\text{CTV})$ for all analysed 4D plan configurations. These overall $V_{95}(\text{CTV})$ errors are calculated by the mean \pm SD of the individual errors given by all three patient geometries combined, with respect to modulated motion scenario (a) A, (b) B, or (c) C, single- or three-field plans, and single scan or five times layered rescanning deliveries.

improve accuracy. However, defining such regions of interest requires a considerable amount of manual work. Therefore, there should be careful consideration about which registration method to choose and whether the additional work is justified. As such, the choice of a slightly less accurate, but more time efficient, DIR method can have advantages as long as the uncertainties resulting from these approaches are understood. Indeed, we would recommend that, given the lack of accuracy demonstrated in this work between the different DIRs, it would make sense to provide error bars on dose calculations that depend on such algorithms. Such error bars could be generated by systematic analysis of dose calculations using different DIR algorithms for different treatment sites, as the one performed on this paper, which is specific for PBS-PT 4D dose calculations from liver cases.

In summary, to understand the influence of a particular registration algorithm on the 4D dose calculation, the accuracy of different DIR methods to estimate the GT results has been analysed. Regarding field-specific geometric registration error, it has been shown that all DIR methods underestimate motion amplitude in the fVOI for all applied fields. This consequently resulted in an

overestimation of the calculated plan index of $V_{95}(\text{CTV})$ for all DIR algorithms in comparison to the GT for most of the 4D plan configurations analysed. The performance of this study is of great importance for the proton therapy community in general, and particularly for PBS-PT, due to its sensitivity to respiratory-induced impacts. As well for passively scattered proton therapy, or even conventional radiotherapy, where DIR is widely used for dose distribution warping and accumulation, we believe that our outcomes, especially on the DIR induced geometric error assessment, still remain noteworthy [33,34].

Conclusion

The demonstrated dosimetric errors induced by different DIR methods indicate the necessity to interpret individual 4D dose distributions for PBS-PT plans for liver cases with caution, and ideally with an error bar. However, by adding fields to the treatment plan and/or using motion mitigation techniques such as rescanning, the impact of DIR motion estimation uncertainties on the 4D dose distributions could be reduced.

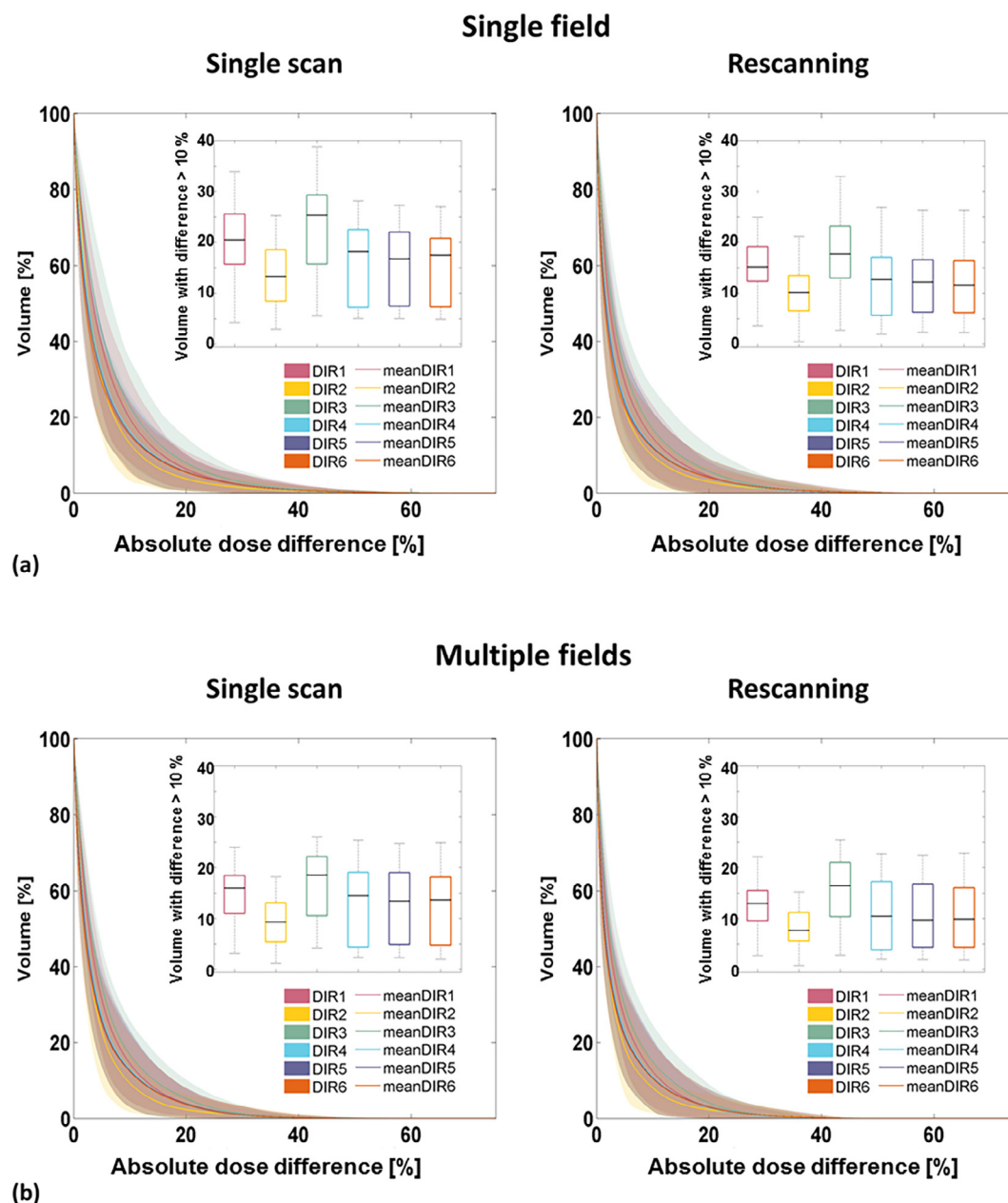


Fig. 5. DDVH band between investigated DIR and GT in the extended CTV (CTV + 1 cm) for all (a) single- vs. (b) three-field 4D plan configurations. Each individual plot gives the information for all motion scenarios (A, B, and C). The boxplots give the statistics of the volume of extended CTV with absolute dose differences higher than 10 %, with respect to DIR method used as input for the 4D dose calculation.

Conflicts of interest

We have no conflicts of interest to disclose.

Acknowledgments

Necessary exported DVF file conversions for DIR5 were kindly supported by Mirada Medical (Oxford, UK). The authors would also like to thank the Computer Vision Laboratory in ETH Zurich (Zurich, Switzerland) for providing scripting functionalities for DIR6 application.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.radonc.2018.03.001>.

References

- [1] Bert C, Grözinger SO, Rietzel E. Quantification of interplay effects of scanned particle beams and moving targets. *Phys Med Biol* 2008;53:2253–65.
- [2] Grassberger C, Dowdell S, Lomax A, Sharp G, Shackelford J, Choi N, et al. Motion interplay as a function of patient parameters and spot size in spot scanning proton therapy for lung cancer. *Int J Radiat Oncol* 2013;86:380–6.
- [3] Dowdell S, Grassberger C, Sharp GC, Paganetti H. Interplay effects in proton scanning for lung: a 4D Monte Carlo study assessing the impact of tumor and beam delivery parameters. *Phys Med Biol* 2013;58:4137–56.
- [4] KØ Noe. Deformable Image Registration for Use in Radiotherapy; 2009.
- [5] Jingu K. Use of deformable image registration for radiotherapy applications. *J Radiol Radiat Ther* 2014;2:1–7.
- [6] Zhang Y, Boye D, Tanner C, Lomax AJ, Knopf A. Respiratory liver motion estimation and its effect on scanned proton beam therapy. *Phys Med Biol* 2012;57:1779–95.
- [7] Kadoya N, Nakajima Y, Saito M, Miyabe Y, Kurooka M, Kito S, et al. Multi-institutional validation study of commercially available deformable image registration software for thoracic images. *Int J Radiat Oncol Biol Phys* 2016;96:422–31.

- [8] Brock KK. Results of a multi-institution deformable registration accuracy study (MIDRAS). *Int J Radiat Oncol Biol Phys* 2010;76:583–96.
- [9] Bender ET, Hardcastle N, Tomé WA. On the dosimetric effect and reduction of inverse consistency and transitivity errors in deformable image registration for dose accumulation. *Med Phys* 2011;39:272–80.
- [10] Salguero FJ, Saleh-Sayah NK, Yan C, Siebers JV. Estimation of three-dimensional intrinsic dosimetric uncertainties resulting from using deformable image registration for dose mapping. *Med Phys* 2010;38:343–53.
- [11] Murphy MJ, Salguero FJ, Siebers JV, Staub D, Vaman C. A method to estimate the effect of deformable image registration uncertainties on daily dose mapping. *Med Phys* 2012;39:573–80.
- [12] Yeo UJ, Taylor ML, Supple JR, Smith RL, Dunn L, Kron T, et al. Is it sensible to “deform” dose? 3D experimental validation of dose-warping. *Med Phys* 2012;39:5065–72.
- [13] Boye D, Lomax T, Knopf A. Mapping motion from 4D-MRI to 3D-CT for use in 4D dose calculations: A technical feasibility study. *Med Phys* 2013;40:61702.
- [14] D Boye. Applications of 4D-MRI in proton therapy; 2016.
- [15] Bernatowicz K, Peroni M, Perrin R, Weber DC, Lomax AJ. 4D dose reconstruction for scanned proton therapy using liver 4DCT-MRI. *Int J Radiat Oncol* 2016;95:216–23.
- [16] Zhang Y, Huth I, Wegner M, Weber DC, Lomax AJ. An evaluation of rescanning technique for liver tumour treatments using a commercial PBS proton therapy system. *Radiother Oncol* 2016;121:281–7.
- [17] Weistrand O, Svensson S. The ANACONDA algorithm for deformable image registration in radiotherapy. *Med Phys* 2015;42:40–53.
- [18] Press release. RaySearch licenses groundbreaking technology from Princess Margaret Hospital; 2011:7–8.
- [19] Samavati N, Velec M, Brock K. A hybrid biomechanical intensity based deformable image registration of lung 4DCT. *Phys Med Biol* 2015;60:3359–73.
- [20] Mirada Medical. Why choose Mirada Registration ? 2002. <http://www.mirada-medical.com/>.
- [21] BD Lucas. Generalized image matching by the method of differences; 1984.
- [22] Vishnevskiy V, Gass T, Szekely G, Tanner C, Goksel O. Isotropic total variation regularization of displacements in parametric image registration. *IEEE Trans Med Imaging* 2017;36:385–95.
- [23] Siddon RL. Fast calculation of the exact radiological path for a three-dimensional CT array. *Med Phys* 1985;12:252–5.
- [24] Lomax A. SFUD, IMPT, and plan robustness. Part Radiother Emerg Technol Treat Cancer 2016:169–94.
- [25] Zenklusen SM, Pedroni E, Meer D. A study on repainting strategies for treating moderately moving targets with proton pencil beam scanning at the new Gantry 2 at PSI. *Phys Med Biol* 2010;55:5103–21.
- [26] Knopf A-C, Hong TS, Lomax A. Scanned proton radiotherapy for mobile targets—the effectiveness of re-scanning in the context of different treatment planning approaches and for different motion characteristics. *Phys Med Biol* 2011;56:7257–71.
- [27] Bernatowicz K, Lomax AJ, Knopf A. Comparative study of layered and volumetric rescanning for different scanning speeds of proton beam in liver patients. *Phys Med Biol* 2013;58:7905–20.
- [28] Rietzel E, Chen GTY, Choi NC, Willet CG. Four-dimensional image-based treatment planning: Target volume segmentation and dose calculation in the presence of respiratory motion. *Int J Radiat Oncol Biol Phys* 2005;61:1535–50.
- [29] Lin L, Kang M, Huang S, Mayer R, Thomas A, Solberg TD, et al. Beam specific planning target volumes incorporating 4DCT for pencil beam scanning proton therapy of thoracic tumors. *J Appl Clin Med Phys* 2015;16:281–92.
- [30] Inoue T, Widder J, van Dijk LV, Takegawa H, Koizumi M, Takashina M, et al. limited impact of setup and range uncertainties, breathing motion, and interplay effects in robustly optimized intensity modulated proton therapy for stage III non-small cell lung cancer. *Int J Radiat Oncol* 2016;96:661–9.
- [31] Y Zhang, F Mueller, DC Weber, AJ Lomax. TH-AB-605-4 dosimetric influence of temporal resolution used by the PBS proton 4D dose calculations. AAPM 59th Annu. Meet. Exhib., Denver, CO, USA; 2017.
- [32] Brock KK, Mutic S, McNutt TR, Li H, Kessler ML. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM Radiation Therapy Committee Task Group No. 132. *Med Phys* 2017;44:e43–76.
- [33] Flampouri S, Jiang SB, Sharp GC, Wolfgang J, Patel AA, Choi NC. Estimation of the delivered patient dose in lung IMRT treatment based on deformable registration of 4D-CT data and Monte Carlo simulations. *Phys Med Biol* 2006;51:2763–79.
- [34] Abe Y, Kadoya N, Arai K, Takayama Y, Kato T, Kimura K, et al. Effect of DIR uncertainty on prostate passive-scattering proton therapy dose accumulation. *Phys Medica* 2017;39:113–20.